

Dkt. 372

PUBLIC VERSION

**UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK**

THE NEW YORK TIMES COMPANY,

Plaintiff,

v.

MICROSOFT CORPORATION, OPENAI,
INC., OPENAI LP, OPENAI GP, LLC,
OPENAI, LLC, OPENAI OPCO LLC,
OPENAI GLOBAL LLC, OAI
CORPORATION, LLC, OPENAI
HOLDINGS, LLC,

Defendants.

Civil Action No. 1:23-cv-11195-SHS

Hon. Sidney H. Stein

**MEMORANDUM OF LAW IN SUPPORT
OPENAI'S OBJECTION TO
NONDISPOSITIVE ORDER PURSUANT
TO FEDERAL RULE OF CIVIL
PROCEDURE 72(a)**

TABLE OF CONTENTS

	<u>Page</u>
I. INTRODUCTION	1
II. FACTUAL ALLEGATIONS	2
III. THE TIMES’S ALLEGATIONS RAISE CORE FAIR USE ISSUES.	3
IV. PROCEDURAL HISTORY.....	5
A. OpenAI seeks highly relevant discovery regarding the New York Times’s alleged market harm.....	5
B. The Times refused to provide fulsome discovery regarding its alleged market harm.	6
C. The Magistrate Judge rejected OpenAI’s request for market-harm discovery without oral argument and without permitting a formal motion to compel.....	7
D. The Magistrate Judge relied on a separate discovery order considering unrelated issues in denying OpenAI’s market-harm requests.....	9
V. STANDARD OF REVIEW	10
VI. ARGUMENT.....	10
A. The Revenue and Traffic Requests seek documents critical to OpenAI’s fair use defense.	11
B. The Licensing Requests seek documents critical to OpenAI’s fair use defense.	15
1. Preexisting Licensing Revenue.....	15
2. “Potential” Licensing Revenue for LLM Training	16
VII. CONCLUSION.....	18

TABLE OF AUTHORITIES

	Page(s)
Federal Cases	
<i>Am. Geophysical Union v. Texaco, Inc.</i> , 60 F.3d 913 (2d Cir. 1994).....	5, 7, 11, 17
<i>Authors Guild, Inc. v. Hathitrust</i> , 755 F.3d 87 (2d Cir. 2014).....	3
<i>Authors Guild v. Google, Inc.</i> , 804 F.3d 202 (2d Cir. 2015).....	4
<i>Baker v. Urb. Outfitters, Inc.</i> , 254 F. Supp. 2d 346 (S.D.N.Y. 2003).....	14
<i>Bill Graham Archives v. Dorling Kindersley, Ltd.</i> 448 F.3d 605 (2d Cir. 2006).....	7, 16
<i>Campbell v. Acuff-Rose Music, Inc.</i> , 510 U.S. 569 (1994).....	3, 5
<i>Castle Rock Ent. v. Carol Publ’g. Grp. Inc.</i> , 150 F.3d 132 (2d Cir. 1998).....	17
<i>Condit v. Dunne</i> , 225 F.R.D. 100 (S.D.N.Y. 2004)	10
<i>Coventry Cap. U.S. LLC v. EEA Life Settlements, Inc.</i> , 333 F.R.D. 60 (S.D.N.Y. 2019)	10
<i>Fitzgerald Publ’g Co. v. Baylor Publ’g Co.</i> , 807 F.2d 1110 (2d Cir. 1986).....	14
<i>Google LLC v. Oracle America, Inc.</i> , 593 U.S. 1 (2021).....	<i>passim</i>
<i>Hachette Book Grp., Inc. v. Internet Archive</i> , 115 F.4th 163 (2d Cir. 2024)	10, 13
<i>Hachette Book Grp., Inc. v. Internet Archive</i> , Case No. 1:20-CV-04160-JGK (S.D.N.Y. 2021) ECF 60	13
<i>Harper & Row Publishers, Inc. v. Nation Enter.</i> , 471 U.S. 539 (1985).....	11, 13

<i>On Davis v. The Gap, Inc.</i> , 246 F.3d 152 (2d Cir. 2001).....	14
<i>R.F.M.A.S., Inc. v. So</i> , 748 F. Supp. 2d 244 (S.D.N.Y. 2010).....	10
<i>SEC v. Citigroup Global Markets, Inc.</i> , 752 F.3d 285 (2d Cir. 2014).....	15
<i>Sony Corp. of Am. v. Universal City Studios, Inc.</i> , 464 U.S. 417 (1984).....	3
<i>Stevens Linen Assocs., Inc. v. Mastercraft Corp.</i> , 656 F.2d 11 (2d Cir. 1981).....	14
<i>Zhulinska v. Niyazov Law Grp., P.C.</i> , Case No. 21-CV-1348 (CBA), 2021 WL 5281115 (E.D.N.Y Nov. 12, 2021).....	12
Federal Statutes	
17 U.S.C. § 107.....	3, 4
Rules	
Fed. R. Civ. P. 26(b)(1).....	10
Fed. R. Civ. P. 72(a)	1, 10, 11, 14
Local Civil Rule 37.2.....	8
Other Authorities	
4 Melville B. Nimmer & David Nimmer, <i>Nimmer on Copyright</i> (2024).....	5
Pierre N. Leval, <i>Toward a Fair Use Standard</i> , 103 Harv. L. Rev. 1105 (1990).....	3, 5

I. INTRODUCTION

Pursuant to Federal Rule of Civil Procedure 72(a), OpenAI¹ seeks to “modify or set aside” two discovery orders denying critical discovery related to OpenAI’s fair use defense. The fourth factor of the fair use test considers the effect of OpenAI’s allegedly infringing use on the market value of the Times’s asserted works. The Times has made sweeping allegations of market harm, claiming that OpenAI has harmed its subscription, licensing, advertising and affiliate revenue, indeed, its very “ability” to provide the service of “[i]ndependent journalism” itself. *See* ECF 170 (“FAC”) ¶¶ 1–2, 5. Consequently, OpenAI sought documents and communications related to: (1) alleged decreases in the Times’s revenue, including the reasons for such decreases; (2) alleged decreases in subscriptions and in traffic to the Times’s website and mobile application, including the reasons for such decreases; and (3) the Times’s licensing efforts, including whether a cognizable licensing market for training large-language models (“LLMs”) exists. The discovery orders erroneously conclude that such evidence is irrelevant.

Indeed, the Times itself has almost entirely conceded the relevance of the discovery in question. Notwithstanding this concession, the Times has either refused to produce altogether or has improperly narrowed the scope of its production in response to the requests at issue, severely hampering OpenAI’s ability to test the Times’s claims of market harm. The orders denying this discovery (the “December Orders”)² fail to address controlling law and rely instead on a prior order that addressed entirely different issues—namely, the Times’s own use and development of

¹ “OpenAI” refers collectively to Defendants OpenAI, Inc.; OpenAI LP; OpenAI GP, LLC; OpenAI, LLC; OpenAI OpCo, LLC; OpenAI Global LLC; OpenAI Corporation LLC; and OpenAI Holdings, LLC.

² *See* ECF 353 (denying OpenAI’s October 22 request for a prehearing conference); ECF 352 (denying November 17 request for a prehearing conference).

generative AI products. *See* ECF 344. Because the December Orders are clearly erroneous and contrary to law, they must be set aside.

The fair use questions at issue in this Objection have significant implications for this case. The December Orders threaten to prejudice the ultimate resolution of this matter by depriving this Court of a complete evidentiary record for summary judgment, trial, and (potentially) appeal. The discovery OpenAI sought is directly relevant under controlling fair-use precedent and falls well within the broad scope of discovery contemplated by Rule 26. OpenAI respectfully requests that the Court intervene and order the production of the market-related discovery to which OpenAI is entitled.

II. FACTUAL ALLEGATIONS

As OpenAI recently explained, ECF 363 at 2, the core allegation in this case is that OpenAI used a “copy of the Internet” to train the LLMs that underlie its ChatGPT service—and that, in so doing, OpenAI created unauthorized copies of “millions of” Times articles published on the Internet. FAC ¶¶ 2, 88. The Times’s central legal theory is that the creation of these copies constituted copyright infringement, rather than fair use.

The Times contends that OpenAI’s alleged use of Times content caused “substantial harm” to the Times in several ways. FAC ¶¶ 5, 154. First, the Times alleges that it has sustained harm to a “well-established market” for licensing its content for commercial purposes, which allegedly generates “millions of dollars” in revenue for the Times each year. *Id.* ¶ 156. Second, the Times alleges that OpenAI products threaten to “divert readers . . . away from” the Times’s website because individuals can access Times content directly through OpenAI’s products, reducing traffic to the Times website and corresponding advertising and subscription revenues. *Id.* ¶ 157. Finally, the Times alleges that OpenAI products threaten affiliate referral revenues

earned by Times sites like Wirecutter by reducing user visits and thus user clicks on paid-by-click affiliate links in Wirecutter articles. *Id.* ¶¶ 128–29, 157.

III. THE TIMES’S ALLEGATIONS RAISE CORE FAIR USE ISSUES

OpenAI intends to present a fair-use defense in this case. “From the infancy of copyright protection, some opportunity for fair use of copyrighted materials has been thought necessary to fulfill copyright’s very purpose, ‘[t]o promote the Progress of Science and useful Arts’” *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 575 (1994). “[C]opyright is ‘not an inevitable, divine, or natural right that confers on authors the absolute ownership of their creations.’” *Authors Guild, Inc. v. Hathitrust*, 755 F.3d 87, 94 (2d Cir. 2014) (quoting Pierre N. Leval, *Toward a Fair Use Standard*, 103 Harv. L. Rev. 1105, 1107 (1990)); see also *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 429–33 (1984) (copyright “has never accorded the copyright owner complete control”). On the contrary, it reflects a pragmatic bargain designed “to stimulate activity and progress in the arts for the intellectual enrichment of the public.” *HathiTrust*, 755 F.3d at 94–95 (internal quotation marks omitted); see also *Sony*, 464 U.S. at 429 (“reward to the [copyright] owner” is “secondary” to the “general benefits derived by the public” (citation omitted)). While copyright holders are granted a “limited monopoly over (and thus the opportunity to profit from) the dissemination of their original works,” *Hathitrust*, 755 F.3d at 95, that monopoly does not come with the right to prevent or demand payment in exchange for fair uses of a work, see *Sony*, 464 U.S. at 432–33; 17 U.S.C. § 107 (fair uses are “not [] infringement[s] of copyright”). In this way, “[t]he fair use doctrine [] permits and requires courts to avoid rigid application of the copyright statute when, on occasion, it would stifle the very creativity which that law is designed to foster.” *Campbell*, 510 U.S. at 577 (cleaned up).

Section 107 of the Copyright Act details four non-exclusive factors to be considered in assessing whether a particular use is “fair”: (1) the purpose and character of the use, including

whether such use is of a commercial nature or is for nonprofit educational purposes; (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and (4) the effect of the use upon the potential market for or value of the copyrighted work. 17 U.S.C. § 107.

With regard to the fourth factor—the effect on the market for the copyrighted work—the analysis must be sensitive not just to whether such an effect exists, but also to its specific nature, its precise magnitude, and whether it relates to “interests that are not protected by the copyright.” *Authors Guild v. Google, Inc.* (“*Google Books*”), 804 F.3d 202, 224 (2d Cir. 2015). Of relevance here, the court “must consider not just the amount but also the source of the [alleged] loss.” *Google LLC v. Oracle Am., Inc.* (“*Oracle*”), 593 U.S. 1, 35 (2021). In *Oracle*, for example, the Supreme Court considered a dispute in which Google was accused of infringing computer code owned by Oracle in its development of the Android smartphone platform. *Id.* at 14–16. In assessing whether the fourth factor supported a finding of fair use, the Court canvassed the evidence in the record relevant to whether Oracle’s alleged losses were attributable to Google’s use of Oracle’s code in the Android platform, ultimately concluding that “the jury could have found that Android did not harm the actual or potential markets for” Oracle’s code. *Id.* at 36. The Court explained that Oracle’s alleged harm—the loss of revenue it might have gleaned from the smartphone market—could not be traced to the copying at issue. It observed that the record indicated that Oracle and its predecessor Sun Microsystems were “poorly positioned to succeed in the mobile phone market,” and that “Sun was beset by business challenges in developing a mobile phone product.” *Id.* The Court further noted that Sun’s CEO had testified that Sun’s failure to build a competing product was not “attributable to Google’s development of Android.” *Id.* at 36. The alleged harm was thus caused by factors other than the alleged infringement. The

Court therefore held that the fourth factor “weighs in favor of fair use.” *Id.* at 40.

The same principles apply when a copyright holder argues that he should have been paid a licensing fee in exchange for a challenged use. While a copyright holder may permit otherwise infringing activity in exchange for a royalty, “not every effect on potential licensing revenues enters the analysis under the fourth factor.” *Am. Geophysical Union v. Texaco, Inc.*, 60 F.3d 913, 929 (2d Cir. 1994). Indeed, “were a court automatically to conclude in every case that potential licensing revenues were impermissibly impaired simply because the secondary user did not pay a fee for the right to engage in the use, the fourth fair use factor would always favor the copyright holder.” *Id.* at 930 n.17.³ On the contrary, “courts have recognized limits on the concept of ‘potential licensing revenues’ by considering only traditional, reasonable, or likely to be developed markets when examining and assessing a secondary use’s ‘effect upon the potential market for or value of the copyrighted work.’” *Id.* (quoting *Campbell*, 510 U.S. at 577). The absence of such a “ready market or means to pay for the use” weighs heavily in favor of fair use. *Id.* at 931.

IV. PROCEDURAL HISTORY

A. OpenAI seeks highly relevant discovery regarding the New York Times’s alleged market harm.

On March 8 and October 21, 2024, OpenAI propounded document requests seeking evidence relevant to market harm and factor four of the fair use inquiry. Of relevance to this Objection, the requests fall roughly into three categories:

³ See also Leval at 1124 (“By definition every fair use involves some loss of royalty revenue because the secondary user has not paid royalties.”); 4 Melville B. Nimmer & David Nimmer, *Nimmer on Copyright* § 13F.08[B] (2024) (“[I]t is always a given that plaintiff suffers a loss of some potential market if that potential is defined as the theoretical market for licensing the very use at bar.”).

- (1) documents and communications related to decreases in the Times’ revenue, including the reasons for such decreases (RFPs 141, 146–48) (“Revenue Requests”);
- (2) documents and communications related to decreases in subscriptions and in traffic to the Times’ platforms, including the reasons for such decreases (RFPs 35, 37, 142–45) (“Traffic Requests”); and
- (3) documents and communications related to the Times’ licensing efforts, including regarding the existence (or non-existence) of a market for LLM training (RFPs 18, 33, 160–61) (“Licensing Requests”).

B. The Times refused to provide fulsome discovery regarding its alleged market harm.

The Times refused to comply with the full scope of OpenAI’s requests. Instead, the Times has only agreed to produce documents that support its *own* position on the fourth fair use factor, while refusing to provide critical evidence relevant to *OpenAI’s* fair use defense. More specifically, in response to the Revenue Requests, the Times agreed only to produce documents sufficient to show revenue generated by the Times’s offerings. It has refused to produce documents or communications related to the *causes* of the alleged decreases in revenue, including analyses and custodial communications attributing these alleged declines in revenue to causes other than OpenAI.⁴ *But see Oracle*, 593 U.S. at 35 (directing courts to consider “not just the amount but also the source of the loss” in analyzing the fourth fair use factor).

In response to the Traffic Requests, the Times made a limited commitment to produce (1) documents sufficient to show trends in readership for the Asserted Works and online

⁴ See ECF 310-1 at 34–36 (NYT Resp. & Obj. to RFPs 146–48) (refusing to produce documents referring to decreases in revenue or “factors that allegedly negatively impacted revenue”).

subscriptions from 2018 through the present; (2) documents sufficient to show trends in traffic generated in connection with content published on the Times’s website from 2018 to the present; and (3) analyses explicitly attributing alleged decreases to OpenAI. But it refuses to produce any communications related to *any other causes* of the alleged declines⁵—exactly the information OpenAI needs in order to demonstrate that any alleged losses are attributable to other causes, not to OpenAI. *See id.*

In response to the Licensing Requests, the Times cabined its production commitment to copies of licenses for the Asserted Works and custodial communications limited to specific attempts to license the Times’s works for LLM training. The Times refuses to produce any communications regarding licensing of the asserted works in its established licensing market, *i.e.*, for purposes other than LLM training. And it refuses despite having specifically alleged the relevance of its established licensing market in its complaint, *see* FAC ¶¶ 50–51, and despite the Second Circuit’s guidance that analyzing the “market for licensing” is essential in the fair use analysis, *Bill Graham Archives v. Dorling Kindersley*, 448 F.3d 605, 614 (2d Cir. 2006). The Times further refuses to produce any communications regarding the existence (or lack thereof) of a functioning market for LLM-training licenses. Such discovery is necessary to assess whether there is a “ready market or means to pay for” the challenged use. *Texaco*, 60 F.3d at 930–31; *see also Oracle*, 593 U.S. at 31–32 (relying on statements of plaintiff’s “executives” in analyzing fair use generally).

C. The Magistrate Judge rejected OpenAI’s request for market-harm discovery

⁵ *See, e.g.*, ECF 310-2 at 27–28 (NYT Resp. & Obj. to RFP 35) (refusing to produce documents relating to the “causes” of “trends in readership or online subscriptions”).

without oral argument and without permitting a formal motion to compel.

Because the requested materials are critical to OpenAI’s fair use defense, OpenAI filed two requests for pre-motion conferences before Magistrate Judge Ona T. Wang to address the Times’s refusal to comply with these requests.⁶ The first request was filed on October 22, 2024, and, as relevant here, sought production as to two Traffic Requests (RFPs 35 and 37). *See* ECF 276. At a status conference on October 30, 2024, OpenAI attempted to argue this initial request, but Judge Wang stated that it would be considered at the next status conference.⁷ Prior to that next conference, OpenAI filed its second request for a pre-motion conference and sought to compel production in response to the Revenue Requests, the Licensing Requests, and the remaining Traffic Requests. *See* ECF 310.

Both requests for pre-motion conferences were slated to be heard at a status conference on December 3, 2024. *See* ECF 350 at 4–5 (joint chart of discovery disputes for discussion at December 3 conference). On December 2, Judge Wang denied both requests via docket entries based entirely on the parties’ three-page pre-motion letters—without conducting a pre-motion conference or allowing OpenAI to file a formal motion to compel. *See* ECF 353 (denying October 22 request);⁸ ECF 352 (denying November 17 request).

⁶ *See* Local Civil Rule 37.2; Judge Wang Individual Practices in Civil Cases § II.B.

⁷ *See* 10/30 Hearing Tr. at 83:9–15 (instructing counsel to “[p]ut that in your chart that’s coming in, in a week anyway” and that the Court would “take a look at that also” before the next conference).

⁸ The Order at ECF 353 also rejected OpenAI’s request as to other RFPs related to the benefits of OpenAI’s services to journalism and the impact of generative AI products on the Times and journalism. *See* ECF 276 at 2–3 (RFPs 70, 76, & 82). The Order as to those requests erred for the same reasons articulated in OpenAI’s separate, previously filed objection. *See* ECF 362 & 363 (Objection to Discovery Order at ECF 344). Rather than repeat those arguments here, OpenAI hereby incorporates them and asks that the Order as to those requests be similarly set aside for the same reasons. *See* ECF 363 at 16–17.

D. The Magistrate Judge relied on a separate discovery order considering unrelated issues in denying OpenAI’s market-harm requests.

Both December Orders stated only that OpenAI’s requests were denied “for the reasons stated in ECF 344,” which was a prior order issued on November 22, 2024 (the “344 Order”), without any further explanation or analysis. In the 344 Order, the Magistrate Judge denied a separate OpenAI request to compel production of documents based on the “public benefits” prong of the fair use defense. The 344 Order addressed the relevance of three issues also critical to the fair use analysis: (1) the Times’s use of third-party artificial intelligence tools, (2) the Times’s use and creation of its own artificial intelligence tools, and (3) the Times’s out-of-court statements about artificial intelligence. ECF 344 at 1. The 344 Order *did not* directly address the types of requests that are at issue in this Objection—the Traffic Requests, Revenue Requests, and Licensing Requests. The 344 Order rested its analysis almost entirely on an incorrect understanding of how the fair use test assesses “public benefits” flowing from any alleged copying of the asserted works. *See id.* at 4–5. Even if that analysis were correct (it is not),⁹ the “public benefits” assessment is simply not relevant to the requests at issue here.¹⁰

Because Supreme Court and Second Circuit cases are uniform in holding that the requested materials are relevant—and indeed, critical—to any fair use defense, OpenAI files this Objection.

⁹ As noted above, OpenAI has separately filed an Objection to the 344 Order. *See* ECF Nos. 362, 363.

¹⁰ Indeed, to the extent the 344 Order has any bearing on the requests at issue here, it underscores that the Times’s refusal to produce was improper. The November Order cites to *Oracle* and the Supreme Court’s acknowledgement that “the source of the loss” is critical when assessing market harm. *See* 344 Order at 4 (*citing Oracle*, 593 U.S. at 35). Yet the December Orders prevent OpenAI from obtaining any party discovery on that issue.

V. STANDARD OF REVIEW

Under Federal Rule of Civil Procedure 72(a), a district court may “modify or set aside” any part of a magistrate judge’s non-dispositive order “that is clearly erroneous or contrary to law.” Fed. R. Civ. P. 72(a). A magistrate judge’s order is “contrary to law” if it “fails to apply or misapplies relevant statutes, case law, or rules of procedure.” *Coventry Cap. U.S. LLC v. EEA Life Settlements, Inc.*, 333 F.R.D. 60, 64 (S.D.N.Y. 2019) (quoting *R.F.M.A.S., Inc. v. So*, 748 F. Supp. 2d 244, 248 (S.D.N.Y. 2010)). A magistrate judge’s order is “clearly erroneous” “if the reviewing court is left with the definite and firm conviction that a mistake has been committed.” *Id.* (quoting *R.F.M.A.S., Inc.*, 748 F. Supp. 2d at 248).

VI. ARGUMENT

OpenAI is entitled to “present evidence” that the Times has not suffered any “market harm to support [its] defense of fair use.” *Hachette Book Grp., Inc. v. Internet Archive*, 115 F.4th 163, 191 (2d Cir. 2024). And OpenAI is entitled under Rule 26 to “obtain discovery regarding *any nonprivileged matter that is relevant* to any party’s claim or defense and proportional to the needs of the case.” Fed. R. Civ. P. 26(b)(1) (emphasis added). “[R]elevance, for purposes of discovery, is an extremely broad concept.” *Condit v. Dunne*, 225 F.R.D. 100, 105 (S.D.N.Y. 2004). Moreover, information sought “need not be admissible in evidence to be discoverable.” Fed. R. Civ. P. 26(b)(1).

The December Orders prevent OpenAI from obtaining discovery that is highly relevant to its defense. Indeed, the Orders do so despite the Times largely conceding that the discovery in question is relevant.¹¹ Supreme Court and Second Circuit cases consistently hold that evidence

¹¹ See ECF 333 (opposing requested discovery on grounds of burden and duplicativeness).

pertaining to the existence or absence of market harm and identifying “the source” of any such harm is not just relevant, but critical to any fair use defense. *Oracle*, 593 U.S. at 35; *Harper & Row Publishers, Inc. v. Nation Enter.*, 471 U.S. 539, 567 (1985) (explaining that a defendant is entitled “to show that [the alleged harm] would have occurred had there been no taking of copyrighted expression”). And common sense dictates that, if the Times’s alleged losses are attributable to competitive pressures unrelated to OpenAI—social media, pandemic-related disruptions, cable news, the proliferation of podcasts—then they cannot represent “market harm” attributable to OpenAI. Even setting aside OpenAI’s fair use defense, such evidence is also critical to challenging the Times’s alleged damages. Lastly, the December Orders also prevent OpenAI from obtaining discovery regarding the alleged licensing market for LLM training, even though the nature of that supposed market—and whether it represents a “traditional, reasonable, or likely to be developed” market—is essential to the fair use analysis. *Texaco*, 60 F.3d at 930–31.

The Orders denying the requested discovery are inconsistent with the scope of discovery under Rule 26 and fail to apply the appropriate case law. Pursuant to Rule 72(a), they must be set aside.

A. The Revenue and Traffic Requests seek documents critical to OpenAI’s fair use defense.

The Revenue and Traffic Requests seek evidence that is central to the fourth factor of OpenAI’s fair use defense. Indeed, the Times largely abandoned any relevance objection to these requests, asserting instead that the requests were unduly burdensome and cumulative.¹² The

¹² The Times broadly conceded relevance before the Magistrate Judge but argued that the requested discovery was unduly burdensome and cumulative. *See* ECF 333. But the Times failed to offer any specific articulation of this alleged burden, failed to propose search terms for the relevant terms, and refused to discuss ways to alleviate this alleged burden while still preserving

Orders denying this discovery, by contrast, did not even address the bases of the Times’s objections, relying instead on a categorical rejection of relevance.

The Revenue and Traffic Requests are all narrowly tailored to the issue of causation. For example, RFPs 142 through 144 target alleged decreases in traffic to the Times’s platforms and, critically, “the reasons for such decreases.”¹³ Similarly, as to the Times’s revenue, RFPs 141 and 146 through 148 target alleged decreases in revenue and “the reasons for such decreases.”¹⁴

Despite conceding relevance as to these requests, the Times refused to produce documents related to the causes of its alleged market harm.¹⁵ But Rule 26 permits discovery into “any nonprivileged matter that is relevant” to OpenAI’s fair use defense, and the case law recounted above makes plain that causation of alleged market harm is critical. Indeed, the Supreme Court in *Oracle* went into great detail analyzing the plaintiff’s alleged losses and whether, among the various factors, Google’s alleged infringement was, in fact, “the source of the loss.” *Oracle*, 593 U.S. at 35. The Court noted that “Sun was beset by business challenges in developing a mobile phone product,” and that Sun’s CEO had testified its losses were not “attributable to Google’s development of Android.” *Id.* at 36. Such evidence—which was pivotal to the Supreme Court’s conclusion that the “fourth factor . . . weighs in favor of fair use”—is

OpenAI’s right to obtain relevant materials. *See id.* A party cannot evade its discovery obligations by offering “merely speculative” concerns regarding burden. *See Zhulinska v. Niyazov Law Grp., P.C.*, Case No. 21-CV-1348 (CBA), 2021 WL 5281115, at *3 (E.D.N.Y. Nov. 12, 2021).

¹³ *See* ECF 310-2 at 32–33.

¹⁴ *See* ECF 310-2 at 31, 34–36.

¹⁵ The Times’s letter brief on this issue also argued that the Traffic and Revenue requests were duplicative of the Times’s other document productions. *See* ECF 333. But none of the cited productions relates to the *causes* of any revenue loss. *See id.*, 333-1 at 2. And as to traffic decreases, the Times has consistently refused to produce the full scope of causation-related documents.

precisely the evidence the Times is refusing to produce. If the Times’s alleged losses derive from factors having nothing to do with OpenAI, then OpenAI is entitled to know. *See also Harper & Row*, 471 U.S. at 567 (holding that a copyright defendant is entitled “to show that [alleged] damage would have occurred had there been no taking of copyrighted expression”); *Hachette*, 115 F.4th at 191 (“[T]he defendant may present evidence of a lack of market harm to support their defense of fair use.”).

The refusal is particularly damaging because the Times has exclusive possession of the granular financial and revenue analysis that would be most probative of this issue. Nothing can substitute for the Times’s own data and analysis regarding the causes of any financial losses it claims to have suffered since OpenAI launched ChatGPT in late 2022. And, indeed, the Second Circuit has rejected attempts to disprove market harm without such direct evidence. For example, in *Hachette*, the defendant had sought market-harm data directly from the plaintiff during district court proceedings. When the plaintiff refused, the defendant filed a pre-motion letter seeking to compel production of “commercial performance data.” *See Hachette Book Grp., Inc. v. Internet Archive*, Case No. 1:20-CV-04160-JGK (S.D.N.Y. 2021), ECF 47 (“IA Letter Brief”). The Magistrate Judge declined to grant the request, *id.* (ECF 60), and the defendant was ultimately forced to rely on expert reports based on third-party data, *Hachette*, 115 F.4th at 191. The Second Circuit rejected these analyses, including by identifying “infirmities” in the use of third-party data for such a critical assessment, and the Court suggested that the defendant should have presented evidence more directly linked to the plaintiff’s alleged harm. *Id.* at 192. But it is precisely such evidence that the defendant had sought—and *been denied*—in discovery in that case. IA Letter Brief at 1–3. OpenAI should not be forced into the same untenable position.

Failure to set aside the December Orders will preclude OpenAI from gathering the evidence that *Hachette* demonstrates is critical, and it will therefore severely prejudice OpenAI’s defense.

Even setting aside the scope of the fair use defense, causation evidence is equally relevant to an assessment of damages. In calculating actual damages, “the primary measure of recovery is the extent to which the market value of the copyrighted work at the time of the infringement has been injured or destroyed *by the infringement*.” *Fitzgerald Publ’g Co. v. Baylor Publ’g Co.*, 807 F.2d 1110, 1118 (2d Cir. 1986) (emphasis added). As noted above, the Times alleges that it has sustained harm in the form of lost revenues for licensing, traffic and advertising, and subscriptions. FAC ¶¶ 156–57. Documents confirming or disproving the Times’s allegations go directly to the assessment of the Times’s actual damages. *See Stevens Linen Assocs., Inc. v. Mastercraft Corp.*, 656 F.2d 11, 15 (2d Cir. 1981) (holding that the measure of copyright damages was the lost profits had there been no infringement). And because “[t]he award of actual damages may not . . . be based upon ‘undue speculation,’” *Baker v. Urb. Outfitters, Inc.*, 254 F. Supp. 2d 346, 356 (S.D.N.Y. 2003) (quoting *On Davis v. The Gap, Inc.*, 246 F.3d 152, 166 (2d Cir. 2001)), not only is information regarding the Times’s alleged harm relevant to its damages theories, it is central to its ability to recover.

Neither the December Orders nor the underlying 344 Order distinguishes or otherwise addresses OpenAI’s arguments. As noted above, the December Orders simply cite to the 344 Order, but they offer no independent analysis tailored to the requests at issue here. And the 344 Order addressed Magistrate Judge Wang’s view of the scope of the “public benefit” analysis, not the causes of the Times’s claimed harm. *See* 344 Order at 4 (assessing scope of the “public benefits” analysis). The 344 Order’s analysis of the “public benefits” prong simply has no bearing on the issues here. *See* FRCP 72(a) (“The district judge in the case must . . . set aside any

part of the order that is clearly erroneous or is contrary to law.”); *SEC v. Citigroup Global Markets, Inc.*, 752 F.3d 285, 297 (2d Cir. 2014) (a lower court’s “failure to make the proper inquiry constitutes legal error”). Moreover, to the extent there is any analytical overlap with the issues in this Objection, the 344 Order supports OpenAI’s request for production. In the midst of its public benefits analysis, the 344 Order explains that discovery relating to a copyright owner’s loss *is* in fact discoverable, and it quotes *Oracle*’s holding that “the source of a loss” is a relevant concern. 344 Order at 4.

Because the December Orders are both contrary to law and clearly erroneous as to the Revenue and Traffic Requests, they must be set aside.

B. The Licensing Requests seek documents critical to OpenAI’s fair use defense.

Evidence related to licensing is relevant in this matter for at least two distinct but related reasons. *First*, the Times alleges that OpenAI’s products have undermined the Times’s preexisting licensing revenues. FAC ¶¶ 5, 157. *Second*, the Times suggests that there exists a *potential* licensing market for use of copyrighted works for LLM training, and that this market was available for a third party wishing to use the Asserted Works for that purpose. *Id.* ¶ 92. OpenAI sought discovery related to both allegations. It requested documents and communications relating to (i) the Times’s efforts to license the asserted works (RFPs 18, 33); (ii) the market for licensing or using the Times’s works for AI training, including documents showing or refuting whether such a market exists (RFP 160); and (iii) whether the Times’s works have ever been used for the purpose of analyzing language patterns (RFP 161).

1. Preexisting Licensing Revenue

With regard to the Times’s preexisting licensing revenue, the requested documents are relevant for the same reasons that the Traffic and Revenue Requests are relevant: OpenAI must be permitted to challenge the Times’s allegations that OpenAI’s products—as opposed to some

other cause—have harmed the Times’s ability to license its works for reproduction, derivative uses, and distribution, among other things. *Bill Graham*, 448 F.3d at 614 (first inquiry under the fourth fair use factor is whether a challenged use “impact[s] [the Times’s] primary market for the sale of [its products]”). While the Times has agreed to produce the licenses themselves, it has refused to produce the associated communications that would put those licenses in context. And it has refused to do so despite simultaneously demanding that OpenAI produce communications related to OpenAI’s licenses. *See* ECF 333, at 3 (Times Letter Brief). OpenAI is producing such licensing communications, and the Times should be required to do the same.

Communications about the Times’s licenses, including negotiation documents themselves, will include pricing and other financial materials that will illuminate the extent of any loss in revenue, and they will also provide context regarding whether OpenAI’s products have any connection to those alleged losses. For example, publicly available information indicates that the Times licenses its content in a variety of markets, including for brand marketing and for journalistic purposes (among others).¹⁶ The relevant markets for these distinct types of licensing are different, as is any theory by which OpenAI’s services could conceivably undermine the Times’s offerings. OpenAI is entitled to discovery as to each of these distinct licensing markets. Production of the licenses alone—which is all the Times has agreed to do—fails to address the underlying details that drive the market harm analysis.

2. “Potential” Licensing Revenue for LLM Training

OpenAI is also entitled to discovery regarding the existence (or lack thereof) of a *potential* market for licensing the Times’s works for LLM training, as the viability and

¹⁶ Compare, e.g., <https://nytlicensing.com/content-brands> (marketing-related licensing), and <https://nytlicensing.com/content-media-organizations> (media-related licensing).

availability of such a market are very much in dispute. The Second Circuit has repeatedly warned that the mere possibility that the defendant could have paid a licensing fee is not a sufficient basis to find market harm. *See, e.g., Texaco*, 60 F.3d at 930–31. Instead, “courts have recognized limits on the concept of ‘potential licensing revenues’ by considering only traditional, reasonable, or likely to be developed markets” *Id.* at 929–30. The absence of such a “ready market or means” to license a work weighs in favor of fair use. *Id.* And the Second Circuit has separately cautioned that, in conducting that analysis, courts must be wary of attempts by copyright owners to engineer the appearance of a workable market in order to “preempt” others from making what would otherwise be fair uses of copyrighted works. *Castle Rock Ent. v. Carol Publ’g. Grp. Inc.*, 150 F.3d 132, 146 n.11 (2d Cir. 1998).

OpenAI is therefore entitled to discovery regarding whether the Times’s licensing efforts related to LLM training reflect a “traditional” or “reasonable” market, or whether such efforts reflect instead an attempt to “preempt” the fair use of its works. *Id.* The Times concedes the relevance of specific licensing efforts, as it has agreed to produce limited communications regarding those efforts. *See* ECF 333 at 2. But it has refused contextual discovery into whether those specific efforts legitimately reflect the existence of a “traditional” or “reasonable” licensing market, or are instead simply the result of a litigation strategy to create the appearance of such a market to rebut OpenAI’s fair use defense. Indeed, documents produced thus far strongly suggest it is the latter. The Times has produced documents reflecting communications with the Copyright Clearance Center in April 2024—during the pendency of this case—in which the Times was urged to [REDACTED]

Ex. A at

NYT_00123603; *see also* Ex. B.¹⁷ The Times cannot be permitted to produce a one-sided version of its licensing efforts in an attempt to [REDACTED] that those efforts reflect a cognizable licensing market, while withholding the critical discovery that demonstrates the opposite. OpenAI is entitled to the discovery necessary to challenge whether this is a legitimate, functional market rather than merely a litigation strategy, and the Licensing Requests are tailored to that effort. If, for example, the Times employees admitted in internal communications that a licensing market for LLMs was not feasible at scale, that the “market” was generated for litigation purposes, or that other impediments foreclosed the viability of such a market, such materials would rebut the Times’s assertions of market harm and strongly support OpenAI’s fair use defense.

VII. CONCLUSION

For the foregoing reasons, OpenAI respectfully requests that this Court vacate the December Orders and order the Times to produce the requested discovery.

KEKER, VAN NEST & PETERS LLP

Dated: December 23, 2024

By: /s/ Andrew F. Dawson

¹⁷ The Times was similarly advised that such licensing efforts would demonstrate [REDACTED]

[REDACTED] Ex. A at NYT_00123603.

ROBERT A. VAN NEST (*pro hac vice*)
R. JAMES SLAUGHTER (*pro hac vice*)
PAVEN MALHOTRA - # 4409397
MICHELLE S. YBARRA (*pro hac vice*)
NICHOLAS S. GOLDBERG (*pro hac vice*)
THOMAS E. GORMAN (*pro hac vice*)
KATIE LYNN JOYCE (*pro hac vice*)
SARAH SALOMON (*pro hac vice*)
633 Battery Street
San Francisco, CA 94111-1809
Telephone: 415 391 5400
Facsimile: 415 397 7188
rvannest@keker.com
rslaughter@keker.com
pmalhotra@keker.com
mybarra@keker.com
ngoldberg@keker.com
tgorman@keker.com
adawson@keker.com
kjoyce@keker.com
ssalomon@keker.com

Attorneys for Defendants OPENAI, INC.,
OPENAI LP, OPENAI GP, LLC, OPENAI,
LLC, OPENAI OPCO LLC, OPENAI
GLOBAL LLC, OAI CORPORATION,
LLC, OPENAI HOLDINGS, LLC

Dated: December 23, 2024

LATHAM & WATKINS LLP

By: /s/ Elana Nightingale Dawson

Andrew M. Gass (*pro hac vice*)

andrew.gass@lw.com

Joseph R. Wetzel

joseph.wetzel@lw.com

505 Montgomery Street, Suite 2000

San Francisco, CA 94111

Telephone: 415.391.0600

Sarang V. Damle

sy.damle@lw.com

Elana Nightingale Dawson (*pro hac vice*)

elana.nightingaledawson@law.com

Michael David (*pro hac vice*)

michael.david@law.com

555 Eleventh Street, NW, Suite 1000

Washington, D.C. 20004

Telephone: 202.637.2200

Allison L. Stillman

alli.stillman@lw.com

Rachel R. Blitzer

rachel.blitzer@lw.com

Herman Yue

herman.yue@lw.com

Luke A. Budiardjo

luke.budiardjo@lw.com

1271 Avenue of the Americas

New York, NY 10020

Telephone: 212.751.4864

Attorneys for OpenAI Defendants

Dated: December 23, 2024

MORRISON & FOERSTER LLP

By: /s/ Rose S. Lee

Joseph C. Gratz (*pro hac vice*)

jgratz@mofo.com

Vera Ranieri (*pro hac vice*)

vranieri@mofo.com

425 Market Street

San Francisco, CA 94105-2482

Telephone: 415.268.7000

Rose S. Lee (*pro hac vice*)

rose.lee@mofo.com

707 Wilshire Boulevard, Suite 6000

Los Angeles, California 90017-3543

Telephone: 213.892.5200

Eric K. Nikolaides

enikolaides@mofo.com

Emily C. Wood

ewood@mofo.com

250 West 55th Street

New York, NY 10019-9601

Telephone: 212.468.8000

Carolyn Homer (*pro hac vice*)

enikolaides@mofo.com

2100 L Street, NW Suite 900

Washington, D.C. 20038

Telephone: 202.887.1500

Attorneys for OpenAI Defendants